

Analysis of a threshold model of social contagion on degree-correlated networks.

Peter Sheridan Dodds^{1,2,*} and Joshua L. Payne^{2,3,†}

¹*Department of Mathematics & Statistics, The University of Vermont, Burlington, VT 05401.*

²*Complex Systems Center & the Vermont Advanced Computing Center,
The University of Vermont, Burlington, VT 05401.*

³*Department of Computer Science The University of Vermont, Burlington, VT 05401.*

(Dated: May 26, 2009)

We analytically determine when a range of abstract social contagion models permit global spreading from a single seed on degree-correlated, undirected random networks. We deduce the expected size of the largest vulnerable component, a network's tinderbox-like critical mass, as well as the probability that infecting a randomly chosen individual seed will trigger global spreading. In the appropriate limits, our results naturally reduce to standard ones for models of disease spreading and to the condition for the existence of a giant component. Recent advances in the distributed, infinite seed case allow us to further determine the final size of global spreading events, when they occur. To provide support for our results, we derive exact expressions for key spreading quantities for a simple yet rich family of random networks with bimodal degree distributions.

PACS numbers: 89.65.-s, 87.19.Xx, 87.23.Ge, 05.45.-a

Over the last decade, the study of real-world, complex networks has grown enormously, fueled in no small part by the advent of readily available, large-scale data sets for real systems [1, 2]. Understanding the coupled dynamics of both the structural evolution of, and processes on, complex networks remains a fertile area of investigation. Of particular importance is the study of contagion, how entities spread through networked systems, as exemplified by the diffusion of practices, beliefs, ideas, and emotions in social networks [3, 4, 5], disease contagion in human and animal populations [6, 7], cascading failures in electrical systems [8, 9], the global spread of computer viruses on the Web [10, 11], and the collapse of financial systems [12].

Our present interest lies in contagion processes where individuals adopt alternate behaviors through imitation of peers. We specifically investigate a threshold model of social contagion on random networks—first proposed and studied by Watts [13]—with the added complication of arbitrary degree-degree correlations. The model's origins lie in the seminal work of Schelling, who employed a threshold model for a population on a checkerboard to gain insight into residential segregation [14, 15]. Granovetter [16] subsequently studied a mean-field, random mixing threshold model which can be seen as a natural limiting case of the model we consider here. Both Schelling and Granovetter's work clearly showed that global uniformity should not be taken to mean that individuals have strong or similar preferences, and that small changes in the distributions of individual preferences could lead to sharp transitions in the system's macroscopic state.

Placing the threshold model on standard random net-

works with arbitrary degree distributions gives rise to a number of novel behaviors not seen in the Granovetter model. For example, individuals are now distinguishable, and a single node changing its state can lead to a complete transition in the entire system's state [13]. Moreover, by greatly limiting nodes' knowledge of the complete network, behaviors that would immediately die out when nodes are aware of all other nodes' states may now spread globally. Balkanization enhances innovation.

We consider binary systems with nodes being in one of two states, σ_0 and σ_1 . Initially, all nodes are in state σ_0 . Our immediate interest is in determining whether or not global spreading is possible when a single node is switched to state σ_1 , and how this condition is affected by altering the level of degree-degree correlations. We then further wish to know two key quantities: (1) the probability global spreading occurs (P_{trig}), and (2) when it does, to what fraction of the entire network (S).

In finding the probability that global spreading takes off after switching a single node to state σ_1 , it is enough for us to view the problem as one of standard percolation [17, 18]. (Determining the final extent of spreading requires a distinct approach but nevertheless still capitalizes on the locally branching nature of random networks [19, 20].) This observation follows from several well known aspects of random networks.

First, infinite, sparse random networks are locally perfectly branching networks, possessing only very long cycles when sufficiently connected. Thus, a node can switch to state σ_1 in the initial stages of a spreading process (started by a single seed) only if a single neighbor switches to σ_1 earlier on. For global spreading to occur, a network must have a percolating component of these easily switched 'vulnerable' nodes [13]. (By percolating component, we mean a connected sub-network containing a non-zero fraction of all nodes.) As the adoption of state σ_1 spreads through this percolating component of vulnerable nodes, non-vulnerable nodes requiring two or more

*Electronic address: peter.dodds@uvm.edu

†Electronic address: joshua.payne@uvm.edu

neighbors in state σ_1 may also begin to switch. However, we need focus only on vulnerable nodes to determine whether global spreading is possible or not. When a percolating vulnerable component exists, it may be viewed as a network's critical mass, and one that is highly susceptible since it possesses its own critical mass—exactly any one of its own members.

Second, our analytic treatment via generating functions is limited to the description of finite network components, which indirectly allows us to describe some aspects of infinite components. Since finite components are pure branching structures (i.e., they contain no cycles of any length), all nodes can only switch to σ_1 due to the conversion of a single neighbor, even in the long run.

We structure the paper as follows. After defining the model fully in Section I, we detail a series of analytic results in Section II, concerning the probability and size of macroscopic spreading events. We confirm, via further analysis and simulations, a number of our calculations for a simple network containing two kinds of nodes in Section III. We offer some concluding remarks in Section IV.

I. MODEL DEFINITION

Each node is initially assigned a ‘response function’ which we take here to be a step function. In effect, each node is given a fixed threshold τ sampled from a distribution $P_{\text{threshold}}(\tau)$. Node states update in synchrony at times $t = 0, 1, 2, \dots$. Each node observes the fraction of its neighbors in state σ_1 , and switches to σ_1 if this fraction meets or exceeds its threshold, τ . Once a node switches to state σ_1 , it remains in state σ_1 permanently (akin to the SI model for disease spreading [6]). Asynchronous updating gives the same results for monotonically increasing response functions.

We define the structure of our networks through edge probabilities, following Newman [21, 22]. In studying a range of structural aspects of degree-correlated networks as well as dynamic phenomena on them (especially contagion processes), it is mathematically convenient to use e_{jk} , the probability that a randomly chosen edge connects a node with degree $j + 1$ to one with degree $k + 1$, rather than j and k . The quantity e_{jk} then refers directly to the number of other edges emanating from the nodes an edge connects. Normalization is uncomplicated, requiring that $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} e_{jk} = 1$.

While this definition of e_{jk} is clear for directed networks, we encounter some subtleties in attempting to accommodate both undirected and directed networks with a single notation (even though many expressions will turn out to be the same). For undirected networks, which are our primary focus here, we evenly divide the probability that a randomly selected edge (now directionless) connects nodes with degree $j + 1$ and $k + 1$ between the quantities e_{jk} and e_{kj} . The chance of a randomly selected edge connecting degree $j + 1$ and $k + 1$ nodes is then $e_{jk} + e_{kj} = 2e_{jk}$, and the matrix formed by the e_{jk}

is symmetric. We are thus effectively retaining a ghost of a directed network, as each link must have a designated first and second node. (For directed networks, e_{jk} need of course not be symmetric, and other complications are possible concerning the correlations between an individual node's in-degree and out-degree.)

We also have the important quantity q_k which is the probability that in randomly choosing an edge, and then randomly choosing one end of that edge, we arrive at a node of degree $k + 1$ (equivalently, the node has k emanating edges). For undirected networks, we have the derivation $q_k = \frac{1}{2} \sum_{j=0}^{\infty} e_{jk} + \frac{1}{2} \sum_{j=0}^{\infty} e_{kj} = \sum_{j=0}^{\infty} e_{jk}$. (The same end expression holds for directed networks, where we must follow the edge's direction.)

The Pearson correlation coefficient r for degree pairs gives us a measure of assortativity, and is given by

$$r = \frac{1}{\sigma_q^2} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} jk(e_{jk} - q_j q_k) \quad (1)$$

where $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$ is the variance in the number of emanating edges from a node arrived at by a random edge. Note that the choice of e_{jk} almost prescribes the form of the resulting network's degree distribution, p_k . The one piece of information missing is the abundance of nodes with no connections, p_0 , which we must define independently. A link between the e_{jk} and p_k for undirected networks follows from the observation that q_k is readily determined from the degree distribution: a randomly chosen edge leads to a degree $k + 1$ node with probability $q_k = (k + 1)p_{k+1}/k_{\text{avg}}$ where $k_{\text{avg}} = \sum_{k=0}^{\infty} k p_k$ is the average degree. We therefore have the connection $q_k = \sum_{j=0}^{\infty} e_{jk} = (k + 1)p_{k+1}/k_{\text{avg}}$. Now, in isolating the p_k , we do not need to know k_{avg} . It is enough to know $p_k \propto q_k/(k + 1)$ for $k > 0$, since we can find the normalization constant (which is in fact k_{avg}) by requiring $\sum_{k=1}^{\infty} p_k = 1 - p_0$.

Finding the probability of triggering a global spreading event reduces to standard percolation once we find the probability that an individual node is vulnerable. We allow this probability to be a function of node degree k , using the notation b_{k1} (more generally, we write the probability that a node of degree k switches to state σ_1 given l contacts in state σ_1 as b_{kl}). We first determine whether nodes are vulnerable or not, removing them from the network in the latter case. Finally, if global spreading occurs on the resulting reduced network, global spreading must occur on the original network.

II. ANALYTIC RESULTS

Building on the work of Newman [21], we find closed form expressions for several probability generating functions related to component sizes. The key probability we need to characterize is $f_{n,j}$, which is the probability that an edge emanating from a degree $j + 1$ node leads to a finite vulnerable subcomponent of size $n = 0, 1, 2, \dots$

Writing the marginal generating function for $f_{n,j}$ as $F_j(x; \vec{b}_1) = \sum_{n=0}^{\infty} f_{n,j} x^n$, we have the following recursive relationship:

$$F_j(x; \vec{b}_1) = x^0 \sum_{k=0}^{\infty} \frac{e_{jk}}{q_j} (1 - b_{k+1,1}) + x \sum_{k=0}^{\infty} \frac{e_{jk}}{q_j} b_{k+1,1} [F_k(x; \vec{b}_1)]^k, \quad (2)$$

where $j = 0, 1, 2, \dots$. In both terms, we have the quantity e_{jk}/q_j which represents the normalized probability that an edge from a degree node $j+1$ leads to a degree $k+1$ node. For undirected networks, we obtain this probability as $(e_{jk} + e_{kj})/(\sum_{k=0}^{\infty} e_{jk} + e_{kj}) = 2e_{jk}/2q_j = e_{jk}/q_j$. The first term in Eq. (2) records the probability that an edge from a degree $j+1$ node leads immediately to a non-vulnerable node, and hence a vulnerable subcomponent of size 0.

The second term involves a composition of generating functions. The generating function F_k is the argument of

$$\sum_{k=0}^{\infty} \frac{e_{jk}}{q_j} b_{k+1,1} x^k, \quad (3)$$

which is itself the generating function for the probability that an edge from a degree $j+1$ node leads to a vulnerable node with k emanating edges [23]. Finally, the x leading the second term in Eq. (2) accounts for the $k+1$ degree node itself.

Our task is now to find critical points indicating the onset of a giant component as we vary network structure by altering the e_{jk} and b_{k1} . We compute the average size of a finite vulnerable component found by following an edge from a degree $j+1$ node, given by $F'_j(1; \vec{b}_1)$. Differentiating Eq. (2), setting $x = 1$, and substituting $q_j = \sum_{k=0}^{\infty} e_{jk}$, we have

$$q_j F'_j(1; \vec{b}_1) = \sum_{k=0}^{\infty} e_{jk} b_{k+1,1} + \sum_{k=0}^{\infty} k e_{jk} b_{k+1,1} F'_k(1; \vec{b}_1). \quad (4)$$

(We have used the fact that $F_k(1; \vec{b}_1) = 1$ for networks without a giant component, and all components contribute to the generating function of $f_{n,k}$.) Rearranging Eq. (4), we have a linear system

$$\sum_{k=0}^{\infty} (\delta_{jk} q_k - k b_{k+1,1} e_{jk}) F'_k(1; \vec{b}_1) = \sum_{k=0}^{\infty} e_{jk} b_{k+1,1} \quad (5)$$

which we write as

$$\mathbf{A}_{\mathbf{E}, \vec{b}_1} \vec{F}'(1; \vec{b}_1) = \mathbf{E} \vec{b}_1 \quad (6)$$

where

$$[\mathbf{A}_{\mathbf{E}, \vec{b}_1}]_{j+1, k+1} = \delta_{jk} q_k - k b_{k+1,1} e_{jk}, \quad (7)$$

$$[\vec{F}'(1; \vec{b}_1)]_{k+1} = F'_k(1; \vec{b}_1), \quad (8)$$

$$[\mathbf{E}]_{j+1, k+1} = e_{jk}, \quad (9)$$

$$\text{and } [\vec{b}_1]_{k+1} = b_{k+1,1} \quad (10)$$

for $j, k = 0, 1, 2, 3, \dots$. A solution exists when $\mathbf{A}_{\mathbf{E}, \vec{b}_1}$ is invertible, i.e., its determinant is non-zero. We then have

$$\vec{F}'(1; \vec{b}_1) = \mathbf{A}_{\mathbf{E}, \vec{b}_1}^{-1} \mathbf{E} \vec{b}_1. \quad (11)$$

Our condition for the onset of global spreading is therefore

$$|\mathbf{A}_{\mathbf{E}, \vec{b}_1}| = 0. \quad (12)$$

with $\mathbf{A}_{\mathbf{E}, \vec{b}_1}$ defined in Eq. (7). We note that for networks where one or more degrees are not present (which we will encounter in our later specific examples), we omit the rows and columns of the above matrices corresponding to those degrees. For simplicity, we present our general results for networks assuming all degrees are represented, observing that adjustments to specific sets of degrees are straightforward.

For uncorrelated networks, upon substituting $e_{jk} = q_j q_k$, the above collapses to the known condition

$$\sum_{k=0}^{\infty} b_{k1} (k-1) \frac{k p_k}{k_{\text{avg}}} = 1, \quad (13)$$

as expected [13]. Moreover, $F'_k(1; \vec{b}_1)$ is seen to be independent of k , since the first node of an edge is now unrelated to the other node and hence also the subcomponent it leads to. We provide details for these calculations in the Appendix.

Returning to the general case of degree-correlated networks, we see that when $b_{k1} = b$, a constant for all k , we have a disease-like contagion process. Furthermore, when $b = 1$, all nodes are vulnerable, and we have found the condition for a giant component, equivalent to that obtained by Newman [21]. (Note that when $b_{k1} = b$, we have $\mathbf{E} \vec{b}_1 = b \vec{q}$ since $[\mathbf{E} \vec{b}_1]_{j+1} = \sum_{k=0}^{\infty} e_{jk} b_{k+1,1} = b \sum_{k=0}^{\infty} e_{jk} = b q_{j+1}$.)

We next consider two probability distributions pertaining to component size: (1) g_n , the probability that a randomly chosen node belongs to a vulnerable component of size n , and (2) h_n , the probability that a randomly chosen node belongs or is adjacent to a vulnerable component of size n . Knowing g_n helps us find the size of the largest vulnerable component, whose presence or absence dictates whether or not global cascades are possible for infinite random networks. The second probability h_n will aid us in determining the probability of triggering a global cascade. The triggering node, which is exogenously

switched to state σ_1 may be either vulnerable and part of the largest vulnerable component, or non-vulnerable and connected to one or more nodes in the largest vulnerable component. For the standard giant component case where $b_{k1} = 1$, we have $g_n = h_n$; otherwise, these distributions are likely distinct.

As for the $f_{k,n}$, we find closed form expressions for the generating functions associated with g_n and h_n . The generating function for g_n satisfies the following relationship:

$$G(x; \vec{b}_1) = x^0 \sum_{k=0}^{\infty} p_k (1 - b_{k1}) + x^1 p_0 b_{01} + x \sum_{k=1}^{\infty} p_k b_{k1} \left[F_{k-1}(x; \vec{b}_1) \right]^k. \quad (14)$$

The x^0 term carries the probability that a randomly selected node will not be in state σ_1 ; the second term accounts for the randomly chosen node being vulnerable but having degree 0; and the third term again uses the composition rule for sums of random variables of random sizes [23]. The generating function $\sum_{k=1}^{\infty} p_k b_{k1} x^k$ corresponds to the probability distribution for a randomly chosen node to have degree k and be vulnerable. Note that the argument F_{k-1} appears in the last term rather than F_k because F_{k-1} by definition corresponded to a degree k node.

The generating function for the triggering distribution satisfies a simplified version of Eq. (14):

$$H(x; \vec{b}_1) = x^1 p_0 + x \sum_{k=1}^{\infty} p_k \left[F_{k-1}(x; \vec{b}_1) \right]^k. \quad (15)$$

For the triggering distribution, the first node is now always switched to state σ_1 , regardless of whether it is itself vulnerable or not. The x^1 term accounts for this initial node having degree 0 and hence being unable to trigger any other node. If the initial node has at least one neighbor, then spreading may occur and we can make use of the basic degree generating function $\sum_{k=1}^{\infty} p_k x^k$ combined with the generating function $F_{k-1}(x; \vec{b}_1)$ for finite vulnerable subcomponent size. As per our previous examples, the preceding factor x in the second term accounts for the initial node. In effect, $b_{k1} = 1$ for the triggering node as it is always forced to be switched to state σ_1 , and indeed setting $b_{k1} = 1$ in Eq. (14) directly yields Eq. (15).

Now, the fraction of nodes in the largest vulnerable component is given by $S_{\text{vuln}} = 1 - G(1; \vec{b}_1)$, since $G(1; \vec{b}_1)$ can be seen as the probability that a random node is part of a finite vulnerable component (including one of size 0). Setting $x = 1$ in Eq. (14), we have

$$S_{\text{vuln}} = 1 - G(1; \vec{b}_1) = \sum_{k=0}^{\infty} p_k b_{k1} - p_0 b_{01} - \sum_{k=1}^{\infty} p_k b_{k1} \left[F_{k-1}(1; \vec{b}_1) \right]^k. \quad (16)$$

In the same fashion as for S_{vuln} , the probability of triggering a cascade can be determined using $H(x; \vec{b}_1)$:

$$S_{\text{trig}} = 1 - H(1; \vec{b}_1) = 1 - \sum_{k=0}^{\infty} p_k \left[F_{k-1}(1; \vec{b}_1) \right]^k. \quad (17)$$

The size of the triggering component can also be obtained by first making the observation that an initially switched node of degree k triggers a cascade with probability

$$S_{\text{trig}}^{(k)} = 1 - \left[F_{k-1}(1; \vec{b}_1) \right]^k. \quad (18)$$

This is the probability that at least one edge from a degree k node leads to a giant component of vulnerable nodes. We then have $S_{\text{trig}} = \sum_{k=0}^{\infty} p_k S_{\text{trig}}^{(k)}$, which is in agreement with Eq. (17).

Both S_{vuln} and S_{trig} depend on $F_k(1; \vec{b}_1)$, for which we obtain a potentially infinite set of closed-form, coupled, nonlinear recursive expressions from Eq. (2):

$$F_j(1; \vec{b}_1) = \sum_{k=0}^{\infty} \frac{e_{jk}}{q_j} (1 - b_{k+1,1}) + \sum_{k=0}^{\infty} \frac{e_{jk}}{q_j} b_{k+1,1} \left[F_k(1; \vec{b}_1) \right]^k \quad (19)$$

where we have substituted $q_j = \sum_{k=0}^{\infty} e_{jk}$. Solving Eq. (19) will almost always involve numerical techniques (though in Section III, we examine a case that has analytic solutions). The uncorrelated, pure random network version for the giant component ($b_{k1} = 1$) may serve as some inspiration. There, we have no dependence on the degree of the initial node and the problem reduces to solving $F(1, \vec{1}) = \sum_{k=0}^{\infty} q_k [F(1, \vec{1})]^k$. Both $F(1, \vec{1}) = 0$ and $F(1, \vec{1}) = 1$ are solutions and any initial estimate in between leads to, upon iteration, an intermediate solution, if one exists. Similarly, iteration of Eq. (19) should generally reach the appropriate fixed point solution.

We can also and more easily determine the average size of all vulnerable components (applicable when no giant vulnerable component is present since we again use $F_k(1; \vec{b}_1) = 1$):

$$G'(1, \vec{b}_1) = \vec{p}^T \vec{b}_1 + k_{\text{avg}} (\mathbf{E} \vec{b}_1)^T \mathbf{A}_{\mathbf{E}, \vec{b}_1}^{-1} (\mathbf{E} \vec{b}_1), \quad (20)$$

where $[\vec{p}]_k = p_k$. If we allow spreading to start from any node by forcing the first node to be in state σ_1 , then the average size is instead given by

$$H'(1, \vec{b}_1) = 1 + k_{\text{avg}} (\mathbf{E} \vec{1})^T \mathbf{A}_{\mathbf{E}, \vec{b}_1}^{-1} (\mathbf{E} \vec{b}_1). \quad (21)$$

We complete our analysis with a description of the expected final size of a global spreading event, when it occurs. We use the results of recent work by Gleeson and Calahane [19], and subsequently Gleeson [20], who have explicated an elegant, general solution to the distributed, infinite seed case for a variety of spreading models on a wide range of random networks.

Their key observation is that a node i (degree k_i) switching to state σ_1 at time step n can only be due to previous switching in nodes within n steps of i (including, trivially, i itself). Furthermore, this neighborhood of i must be a pure branching network, i.e., a tree rooted at i . At time step $n = 1$, only i 's immediate neighbors influence i , and these are in state σ_1 with probability ϕ_0 . We write the probability that one of i 's edges connects to a node in state σ_1 at time step $n + 1$ as $\theta_{k_i, n}$. At time step $n = 2$, the effect of next neighbors is felt, and i 's degree k neighbors are now in state σ_1 with probability $\theta_{k, 1}$ which in turn depends generally on $\theta_{k', 0} = \phi_0$, $k' \geq 1$. Allowing n to increase, we are lead to a recursive expression for the $\theta_{k, n}$ as follows. The expected size of a global spreading event given a fraction ϕ_0 of nodes active at time $t = 0$ is ϕ_∞ , which is obtained in the $n \rightarrow \infty$ limit of the following equations [20]:

$$\phi_{n+1} = \phi_0 + (1 - \phi_0) \sum_{k=0}^{\infty} p_k \sum_{i=0}^k \binom{k}{i} \theta_{k, n}^i (1 - \theta_{k, n})^{k-i} b_{ki} \quad (22)$$

where

$$\theta_{j, n+1} = \phi_0 + (1 - \phi_0) \times \sum_{k=1}^{\infty} \frac{e_{j-1, k-1}}{q_{j-1}} \sum_{i=0}^{k-1} \binom{k-1}{i} \theta_{k, n}^i (1 - \theta_{k, n})^{k-1-i} b_{ki} \quad (23)$$

and, again, $\theta_{k, 0} = \phi_0$ and b_{ki} is the probability that a degree k node switches to state σ_1 if exactly i of its neighbors have switched.

We thus observe a symmetry between the explanations for when global spreading may occur from a single seed and if so, to what extent. The former concerns the progress of a spreading event as it moves outward from a single seed through a random branching network, and the latter hinges on how switching converges on a central node, again traversing a branching network, but in the reverse direction (this process is similarly described in [24] for a particular disease spreading model involving repeated contacts).

Note that ϕ_n must always increase or stay the same, since nodes never switch off, and must therefore approach a limit as $n \rightarrow \infty$. This implies that node i feels the effect of initial activations within a finite number of steps, and indeed we see that the approach to ϕ_∞ is rapid. Whether or not a small seed takes off can be determined from Eq. (23) by examining the matrix $d\theta_{j, n+1}/d\theta_{k, n}$ evaluated at $\theta_{k, n} = 0$ [20], and yields a condition equivalent to our requirement $|\mathbf{A}| = 0$ (Eq. (12)).

While these equations are derived with the assumption of an infinite seed (if an arbitrarily small one, fraction-wise), they nevertheless perform extremely well in the limit $\phi_0 \rightarrow 0$. Even so, they cannot capture the salience of a global vulnerable component for the single seed case, which must be present for spreading to succeed. By comparison, for the infinite seed model, spreading always occurs if the cascade condition is met. However, as noted

in [20], by separately determining S_{trig} , we can account for the major features of the single seed model: the probability of generating a global spreading event, S_{trig} , and its expected size ϕ_∞ . We note that in [20], an expression for S_{trig} for uncorrelated random networks was obtained using θ_∞ ; here, we have generalized S_{trig} to uncorrelated networks directly from the standard generating function approach.

III. APPLICATION TO A SIMPLE FAMILY OF RANDOM NETWORKS

For a tractable test case, we examine a family of infinite random networks with nodes having either of two degrees k_a and k_b where $k_a < k_b$. We consider the general class of threshold profiles \vec{b}_1 such that degree k_a nodes are vulnerable and degree k_b nodes are not. We are able to obtain expressions for the size of the vulnerable and triggering components as a function of assortativity, and we make comparisons with simulations for the $(k_a, k_b) = (3, 4)$ case.

We wish to consider the full range of assortativity and this constrains the relative probabilities of the two degrees in the following way. We specifically need

$$p_{k_a} = \frac{k_b}{k_a} p_{k_b}, \quad (24)$$

since when $r = -1$, degree k_a nodes connect only to degree k_b nodes, and so there must be relatively fewer of the latter. We therefore have

$$p_{k_a} = \frac{k_b}{k_a + k_b}, \quad p_{k_b} = \frac{k_a}{k_a + k_b}, \quad \text{and } p_k = 0 \text{ otherwise.} \quad (25)$$

Values of r other than -1 add no further constraints. The corresponding \mathbf{E} matrix is independent of k_a and k_b and has the form

$$\mathbf{E} = \frac{1}{4} \begin{bmatrix} 1+r & 1-r \\ 1-r & 1+r \end{bmatrix}, \quad (26)$$

where we have left out irrelevant rows and columns of zeros. The condition for a giant component given by Eqs. (7) and (12) is then

$$0 = |\mathbf{A}_{\mathbf{E}, \vec{b}_1}| = \frac{1}{4} \begin{vmatrix} 1 - \frac{1}{2}(k_a - 1)(1+r) & 0 \\ \frac{1}{2}(k_a - 1)(r-1) & 1 \end{vmatrix} \quad (27)$$

$$= \frac{1}{2} - \frac{1}{4}(k_a - 1)(1+r).$$

We therefore have a phase transition when

$$r = -1 + \frac{2}{k_a - 1}, \quad (28)$$

with a giant component of vulnerable nodes arising and then growing as assortativity increases. (Note that when no giant component exists, $|\mathbf{A}|$ is positive.) For the case $k_a = 3$, the phase transition occurs at $r = 0$, and as k_a

increases, the location of the phase transition moves to more negative values of r .

We next determine the size of the vulnerable and triggering components. We find the probabilities of reaching a finite component from degree k_a and k_b nodes via Eq. (19):

$$F_{k_a-1}(1; \vec{b}_1) = \frac{1-r}{2} + \frac{1+r}{2} \left[F_{k_a-1}(1; \vec{b}_1) \right]^{k_a-1}, \quad (29)$$

$$F_{k_b-1}(1; \vec{b}_1) = \frac{1+r}{2} + \frac{1-r}{2} \left[F_{k_a-1}(1; \vec{b}_1) \right]^{k_a-1}. \quad (30)$$

Only the first equation need be solved and is in general a $k_a - 1$ degree polynomial in $F_{k_a-1}(1; \vec{b}_1)$. We now specialize our results for the case $k_a = 3$. We have a quadratic equation from which we obtain

$$F_2(1; \vec{b}_1) = \begin{cases} 1 & \text{for } r \leq 0, \\ \frac{1-r}{1+r} & \text{for } r \geq 0. \end{cases} \quad (31)$$

and

$$F_{k_b-1}(1; \vec{b}_1) = \begin{cases} 1 & \text{for } r \leq 0, \\ \frac{1+r}{2} \left[1 + \left(\frac{1-r}{1+r} \right)^3 \right] & \text{for } r \geq 0. \end{cases} \quad (32)$$

Note that in finding $F_2(1; \vec{b}_1)$, we have had to make a choice between the two roots of the quadratic given in Eq. (31). First, we have only one realistic solution in the interval $[0, 1]$ for $r < 0$, namely $F_2(1; \vec{b}_1) = 1$ (the other solution, $(1-r)/(1+r)$, exceeds 1 for $r < 0$). Second, in considering the limit $r = 1$ where degree 3 nodes are connected only to other degree 3 nodes forming a giant component, we require that the probability of reaching a finite component must drop to 0. We therefore see that $(1-r)/(1+r)$ must be the correct choice since then $F_2(1; \vec{b}_1) = 0$ at $r = 1$. We further observe that the two roots cross when $r = 0$, and for $r > 0$, the root $(1-r)/(1+r)$ is less than the root at unity. Thus, by an expectation of continuity, $(1-r)/(1+r)$ must be the solution for all of $r \geq 0$.

We next determine the size of the vulnerable component and triggering probabilities as functions of r using Eqs. (16), (17), and (18). For $r < 0$, we have $S_{\text{vuln}} = S_{\text{trig}} = 0$, and for $r \geq 0$,

$$S_{\text{vuln}} = \frac{k_b}{k_b + 3} \left[1 - \left(\frac{1-r}{1+r} \right)^3 \right], \quad (33)$$

$$S_{\text{trig}}^{(3)} = 1 - \left(\frac{1-r}{1+r} \right)^3, \quad (34)$$

$$S_{\text{trig}}^{(k_b)} = 1 - \left(\frac{1+r}{2} \right)^{k_b} \left[1 + \left(\frac{1-r}{1+r} \right)^3 \right]^{k_b},$$

and

$$S_{\text{trig}} = 1 - \frac{k_b}{k_b + 3} \left(\frac{1-r}{1+r} \right)^3 - \frac{3}{k_b + 3} \left(\frac{1+r}{2} \right)^{k_b} \left[1 + \left(\frac{1-r}{1+r} \right)^3 \right]^{k_b}. \quad (35)$$

As shown in Fig. 1(a), we obtain excellent agreement between equations and the output of simulations for $k_b = 4$ on networks with $N = 0.98 \times 10^5$ nodes and $\tau = 1/3$. We used this particular value of N since it is the nearest number to 10^5 that is divisible by 7 [required by Eqs. (24) and (25)] and, for convenience, also a multiple of 100. As was observed for the original threshold model on random networks [13], we find that well-connected networks near the transition to a non-percolating largest vulnerable component are of a ‘robust-yet-fragile’ nature [26]. We see in Fig. 1(a) that if assortativity r is just slightly positive, and if the initial seed is in the small triggering component present, then spreading reaches the entire network. These otherwise highly resilient networks have an Achilles heel that leads to a complete transition in individual node states. Furthermore, finite size effects may be significant: for the networks with $N = 9800$ nodes, our simulations show that global spreading is possible for r as low as -0.07 .

While our expressions for S_{vuln} and S_{trig} only depend coarsely on \vec{b}_1 , the final extent of a global spreading event S is more sensitive to changes in node response functions. Taking as an example one of our bimodal networks with $k_a = 3$ and $k_b = 10$, we see any value of τ in $(\frac{1}{10}, \frac{1}{3}]$ is equivalent to $\tau = \frac{1}{3}$, as far as S_{vuln} and S_{trig} are concerned, since only degree 3 nodes are vulnerable. However, whether or not degree 10 nodes are vulnerable to i neighbors switching to state σ_1 depends on τ meeting and exceeding $\frac{i}{10}$. For the case $k_a = 3$ and $k_b = 4$, the results we present here apply for all τ in $(\frac{1}{4}, \frac{1}{3}]$.

For the final size of global spreading events, S , manipulations of Eq. (23) for $k_a = 3$ and $k_b = 4$ show we need to solve the following for $\theta_{3,\infty}$:

$$\theta_{3,\infty} = \frac{1+r}{2} i (2 - \theta_{3,\infty}) \theta_{3,\infty} + \frac{1-r}{2} (3 - \theta_{4,\infty}) \theta_{4,\infty}^2 \quad (36)$$

where

$$\theta_{4,\infty} = \frac{1}{1-r} \theta_{3,\infty} (1 - 3r + 2r\theta_{3,\infty}) \quad (37)$$

for $0 \leq r < 1$. Knowing that $\theta_{3,\infty} = 1$ is a solution of Eq. (36) means we can reduce the problem to solving for the roots of a quartic, for which we could obtain a complete analytic solution. For our purposes, numerical simulation is sufficient.

Our simulations and numerical analyses indicate the presence of another kind of phase transition in S . At the upper limit of $r = 1$, the network is separated into two giant, fully-connected components, one comprising solely

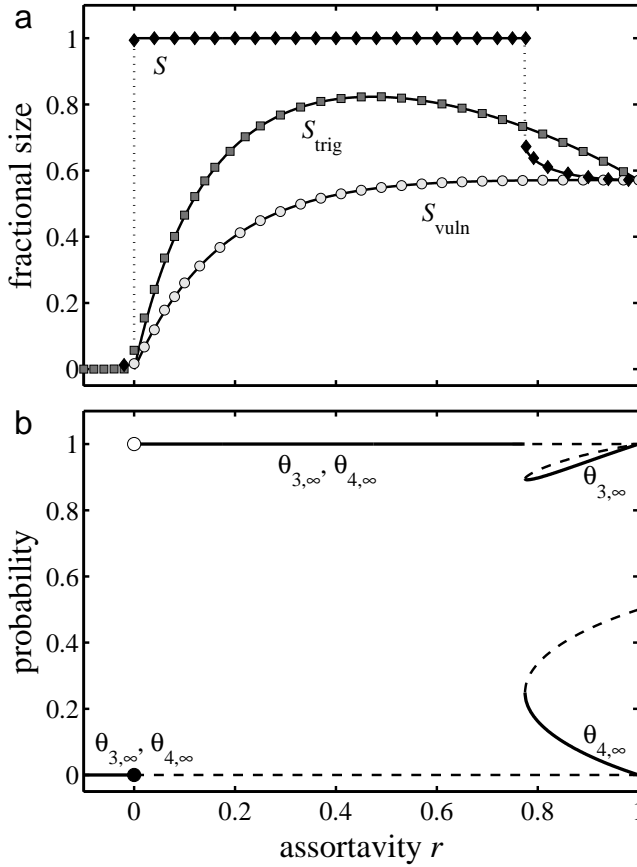


FIG. 1: **(a)** Three key spreading measures as a function of assortativity r for networks with $P_k = \frac{4}{7}\delta_{k3} + \frac{3}{7}\delta_{k4}$. Solid lines indicate theoretical curves and symbols represent measurements from simulations. The three curves are (1) the fractional size of the largest vulnerable component, S_{vuln} , (Eq. (33), circles); (2) the fractional size of the largest triggering component, S_{trig} , (Eq. (35), squares); and (3) the fractional size of global spreading events S (Eq. (22), diamonds). The upper discontinuous phase transition in S occurs at $r = r_c^{\text{upper}} = 0.77485 \pm 0.00001$. The data shown were obtained from networks with $N=0.98 \times 10^5$ nodes (correlations were generated using a shuffling algorithm described in [25]) with initial seeds placed at each node of 10 sample networks for a total of 9.8×10^5 samples per each value of r . **(b)** Solutions for $\theta_{3,\infty}$ and $\theta_{4,\infty}$, the probability that edges leading to degree 3 and 4 nodes will be from nodes in state σ_1 in the long time limit (see Eqs. (36) and (37).) Solid and dashed lines indicate stable and unstable ones respectively. When r reaches r_c^{upper} , a repeated solution appears for $(\theta_{3,\infty}, \theta_{4,\infty})$, distinct from $(1, 1)$, giving rise to the discontinuous upper phase transition in S .

$k = 3$ degree nodes and the other $k = 4$ degree nodes (the giant component is of fractional size 1 for all other values of r). Spreading therefore occurs only in the $k = 3$ degree node component when $r = 1$, and, as shown in Fig. 1(a), the three quantities S , S_{vuln} , and S_{trig} all equal $4/7$ for complete assortativity.

As we decrease r however, the possibility that a degree

$k = 4$ node is connected to two (or more) degree $k = 3$ nodes in the giant vulnerable component increases. The expected size ϕ_∞ of a global spreading event grows gradually until it reaches a discontinuous phase transition and ϕ_∞ jumps to 1. Through numerical analysis of Eq. (36), we see that as r increases, a repeated root $\theta_{3,\infty}$ appears when $r = r_c^{\text{upper}} \simeq 0.775$. The roots become real and separate with both of them limiting to 1 as $r \rightarrow 1$. Fig. 1(b) shows how both $\theta_{3,\infty}$ and $\theta_{4,\infty}$ behave as a function of r . Note that the discontinuity in the final size of a global spreading event is not reflected in the curves for S_{trig} or S_{vuln} , which both change smoothly for $0 \leq r \leq 1$.

IV. CONCLUDING REMARKS

We have developed a series of theoretical expressions for a general class of spreading models on random networks with tunable degree assortativity r . Our main results are the derivation of the fractional size of the giant vulnerable and triggering components as a function of degree-degree distribution e_{jk} . When we allow all nodes to be susceptible with uniform probability b , our results reduce to the standard disease-like spreading case, and when we further set $b = 1$, we obtain the conditions for the existence of a giant component.

When we set $r = 0$, we also retrieve the condition for the existence of a giant vulnerable component (Eq. (12)) along with its size (Eq. (16)) in pure random networks with arbitrary degree distributions [13]. For the triggering component, aside from our work for general r , we have also added to known results for the $r = 0$ case. We have obtained the probability that an exogenously activated node of degree k may trigger a global spreading event (Eq. (18)) and hence the probability that a randomly chosen node may do the same (Eq. (17)). Our work complements the results of Gleeson [20] which determines the final size of a global spreading event (Eqs. (22) and (23)). We note that a basic description of many kinds of spreading from a single initial seed must separately report (1) the probability of a global spreading event P_{trig} , and (2) the expected final size S . The distribution of final sizes is often bimodal and an overall expected size conflates P_{trig} and S in a potentially misleading way, since no size other than 0 and S are in fact possible. Finally, the simple family of random networks we have considered here with $P_k = \frac{4}{7}\delta_{k3} + \frac{3}{7}\delta_{k4}$ have allowed us to demonstrate, for one example, the validity of our theoretical results for $S_{\text{vuln}}(r)$ and $S_{\text{trig}}(r)$.

Acknowledgments

JLP was supported by a Graduate Research Fellowship awarded by Vermont EPSCoR (NSF EPS 0701410). The authors are grateful for the computational resources provided by the Vermont Advanced Computing Center which is supported by NASA (NNX 08A096G).

APPENDIX: CALCULATIONS REGARDING COLLAPSE OF GENERAL GLOBAL SPREADING CONDITION FOR UNCORRELATED NETWORKS

In the main text, we noted that the general condition for global spreading, that $|\mathbf{A}_{\mathbf{E}, \vec{b}_1}| = 0$ (Eq. (12)), collapses to the appropriate condition for uncorrelated networks when we set $e_{jk} = q_j q_k$ (Eq. (13)). We demonstrate this in two distinct ways. First, we take a direct route by setting $e_{jk} = q_j q_k$ in the definition of $\mathbf{A}_{\mathbf{E}, \vec{b}_1}$:

$$\begin{aligned} [\mathbf{A}_{\mathbf{E}, \vec{b}_1}]_{j+1, k+1} &= \delta_{jk} q_k - k b_{k+1,1} q_j q_k \\ &= (\delta_{jk} - q_j k b_{k+1,1}) q_k. \end{aligned} \quad (38)$$

The q_k that has been factored out stands as a multiple for the k th column, and therefore, by multilinearity of determinants [27], contributes to the determinant of $\mathbf{A}_{\mathbf{E}, \vec{b}_1}$ as a simple overall multiplicative factor. In finding $|\mathbf{A}_{\mathbf{E}, \vec{b}_1}|$, we can therefore focus on the matrix $[\delta_{jk} - q_j k b_{k+1,1}]$. We find the determinant of this matrix by finding its eigenvalues, which are the eigenvalues of $[-q_j k b_{k+1,1}]$ incremented by 1 due to the addition of the identity matrix. Since $[-q_j k b_{k+1,1}]$ is an outer product of the two vectors $[-q_j]$ and $[k b_{k+1,1}]$, we have a rank one matrix which therefore has a sole non-zero eigenvalue corresponding to an eigenvector in the direction of $[-q_j]$. We find the non-zero eigenvalue by explicitly applying the matrix $[-q_j k b_{k+1,1}]$ to $[-q_j]$, giving us $\lambda_1 = -\sum_{k=0}^{\infty} k b_{k+1,1} q_k$. Shifting the summation index, we have $\lambda_1 = -\sum_{k=1}^{\infty} (k-1) b_{k,1} q_{k-1}$, and $\lambda_i = 0$ for $i \geq 2$. We now add 1 to all eigenvalues to obtain $\lambda'_1 = 1 - \sum_{k=1}^{\infty} (k-1) b_{k,1} q_{k-1}$ and $\lambda'_i = 1$ for $i \geq 2$. The determinant of $\mathbf{A}_{\mathbf{E}, \vec{b}_1}$ is given by the product of the λ'_i along with the product of the q_k elements we initially factored out:

$$|\mathbf{A}_{\mathbf{E}, \vec{b}_1}| = \left(\prod_{k=0}^{\infty} q_k \right) \left(1 - \sum_{k=1}^{\infty} (k-1) b_{k,1} q_{k-1} \right). \quad (39)$$

Since the q_k are not equal to 0, the condition that $|\mathbf{A}_{\mathbf{E}, \vec{b}_1}| = 0$ for uncorrelated networks depends on the

second term: $0 = 1 - \sum_{k=1}^{\infty} b_{k,1} (k-1) q_{k-1}$. Substituting $q_k = k p_k / k_{\text{avg}}$ gives us the desired condition of Eq. (13).

Secondly, we can also show that $F'_k(1; \vec{b}_1)$ is independent of k , meaning that the expected size of a finite subcomponent found at the other end of an edge connected to a node of degree $k+1$ does not depend on the latter's degree. This will also lead us to the same condition for global spreading we found above. To do so, we return to Eq. (6) and set $e_{jk} = q_j q_k$:

$$\sum_{k=0}^{\infty} (\delta_{jk} q_k - k b_{k+1,1} q_j q_k) F'_k(1; \vec{b}_1) = \sum_{k=0}^{\infty} q_j q_k b_{k+1,1}. \quad (40)$$

Breaking the left hand side into two pieces, carrying out the first summation, and then dividing through by q_j , we obtain

$$F'_j(1; \vec{b}_1) - \sum_{k=0}^{\infty} k b_{k+1,1} q_k F'_k(1; \vec{b}_1) = \sum_{k=0}^{\infty} q_k b_{k+1,1}. \quad (41)$$

Upon rearranging, we can now express $F'_j(1; \vec{b}_1)$ as a function of terms not involving j (even if they include summations involving $F'_k(1; \vec{b}_1)$), thereby demonstrating independence:

$$F'_j(1; \vec{b}_1) = \sum_{k=0}^{\infty} k b_{k+1,1} q_k F'_k(1; \vec{b}_1) + \sum_{k=0}^{\infty} q_k b_{k+1,1}. \quad (42)$$

Setting $F'_k(1; \vec{b}_1) = F'(1; \vec{b}_1)$ for all k , the above is now solvable, and we find

$$F'(1; \vec{b}_1) = \frac{\sum_{k=0}^{\infty} q_k b_{k+1,1}}{1 - \sum_{k=0}^{\infty} k b_{k+1,1} q_k}. \quad (43)$$

Finally, we see that setting the denominator to 0 recovers the condition for global spreading in uncorrelated networks.

-
- [1] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
 - [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Physics Reports* **424**, 175 (2006).
 - [3] J. Coleman, H. Menzel, and E. Katz, *Medical Innovations: A Diffusion Study* (Bobbs Merrill, New York, 1966).
 - [4] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, *Emotional Contagion*, Studies in Emotion and Social Interaction (Cambridge University Press, Cambridge, UK, 1993).
 - [5] J. H. Fowler and N. A. Christakis, *BMJ* **337**, article #2338 (2008).

- [6] J. D. Murray, *Mathematical Biology* (Springer, New York, 2002), Third ed.
- [7] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
- [8] M. L. Sachtjen, B. A. Carreras, and V. E. Lynch, *Phys. Rev. E* **61**, 4877 (2000).
- [9] R. Kinney, P. Crucitti, R. Albert, and V. Latora, *Eur. Phys. J. B* **46**, 101 (2005).
- [10] M. E. J. Newman, S. Forrest, and J. Balthrop, *Phys. Rev. E* **66**, 035101 (2002).

- [11] J. Balthrop, S. Forrest, M. Newman, and M. W. Williamson, *Science* **304**, 527 (2004).
- [12] D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems* (Princeton University Press, Princeton, NJ, 2003).
- [13] D. J. Watts, *Proc. Natl. Acad. Sci.* **99**, 5766 (2002).
- [14] T. Schelling, *J. Math. Sociol.* **1**, 143 (1971).
- [15] T. C. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).
- [16] M. Granovetter, *Am. J. Sociol.* **83**, 1420 (1978).
- [17] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, Washington, D.C., 1992), Second ed.
- [18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [19] J. P. Gleeson and D. J. Cahalane, *Phys. Rev. E* **75**, 056103 (2007).
- [20] J. P. Gleeson, *Phys. Rev. E* **77**, 046117 (2008).
- [21] M. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [22] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
- [23] The composition rule applies for a sum of identical random variables where the number summed is itself a random variable (the emanating degree in our particular case). Specifically, if we have two non-negative integer variables $N^{(1)}$ and $N^{(2)}$ with probabilities $P_{N^{(1)}}^{(1)}$ and $P_{N^{(2)}}^{(2)}$, and generate a third random variable as $N^{(3)} = \sum_{l=1}^{N^{(2)}} N^{(1)}_l$, then $\theta^{(3)}(x; \vec{b}_1) = \theta^{(2)}(\theta^{(1)}(x; \vec{b}_1))$ where $\theta^{(i)}$ is the generating function for $P_{N^{(i)}}^{(k)}$.
- [24] O. Diekmann, M. C. M. De Jong, and J. A. J. Metz, *J. Appl. Prob.* **35**, 448 (1998).
- [25] J. L. Payne and M. J. Eppstein, in *Proceedings of the Genetic and Evolutionary Computation Conference*, edited by M. Keijzer *et al.* (2008), p. 241.
- [26] J. Carlson and J. Doyle, *Phys. Rev. Lett.* **84**, 2529 (2000).
- [27] G. Strang, *Introduction to Linear Algebra* (Cambridge Wellesley Press, Wellesley, MA, 2009), 4th ed.